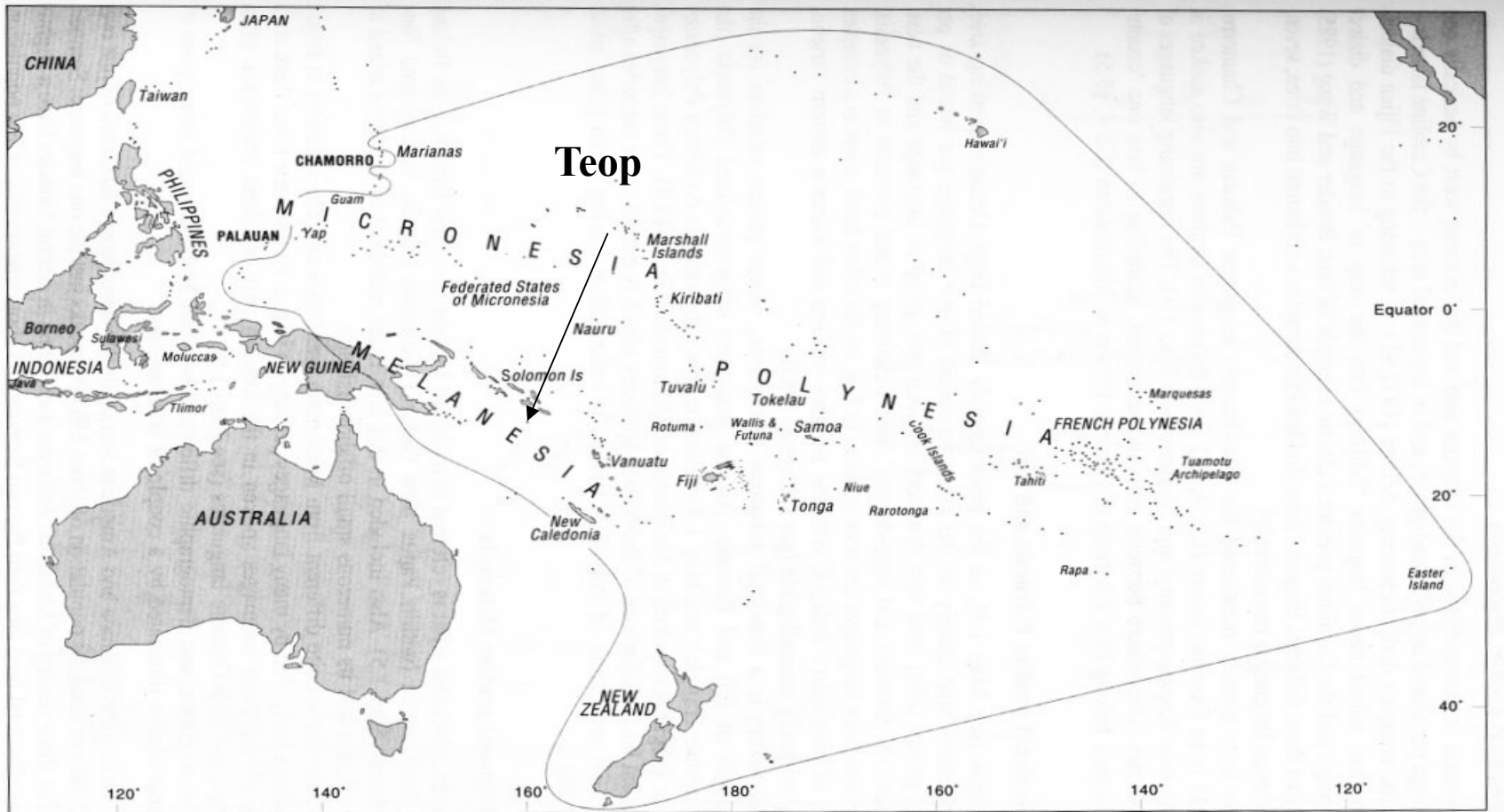


Turning a linguist's lexical data base into a community dictionary
 Ulrike Mosel, University Kiel

International Conference on Language Documentation and Conversation
 March 2009

The Teop language documentation project



MAP 1.2 THE BOUNDARIES OF THE OCEANIC SUBGROUP

Language documentation

Corpus

Recordings
with transcriptions,
translations,
comments,
pictures

Edited versions
of recordings
written texts

SKETCH GRAMMAR

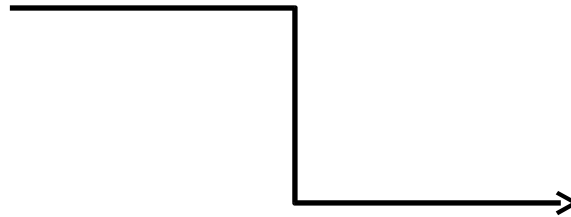
examples

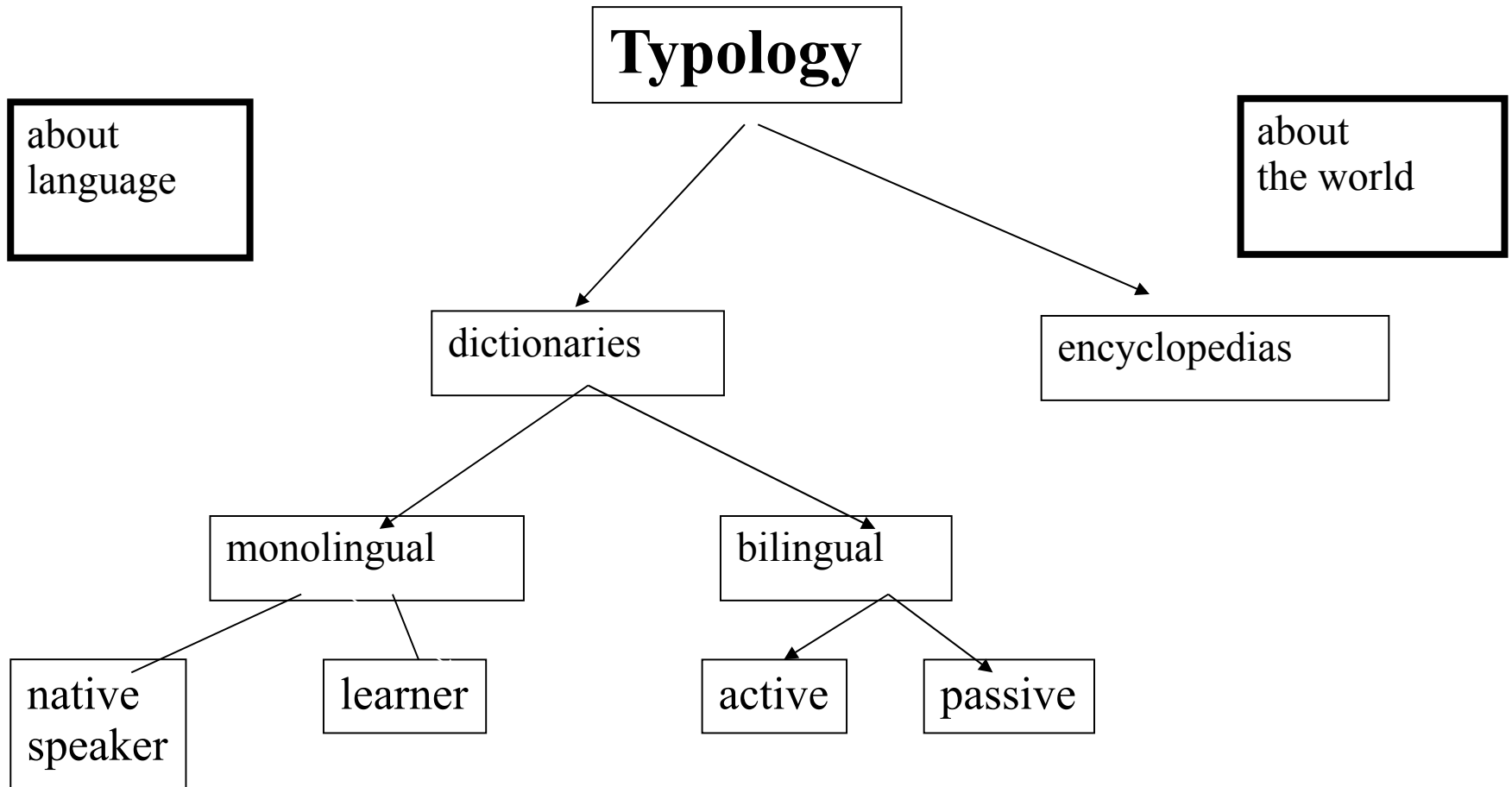
LEXICAL DATABASE

headwords
translations/
examples
Beispiele
Belegstellenangabe

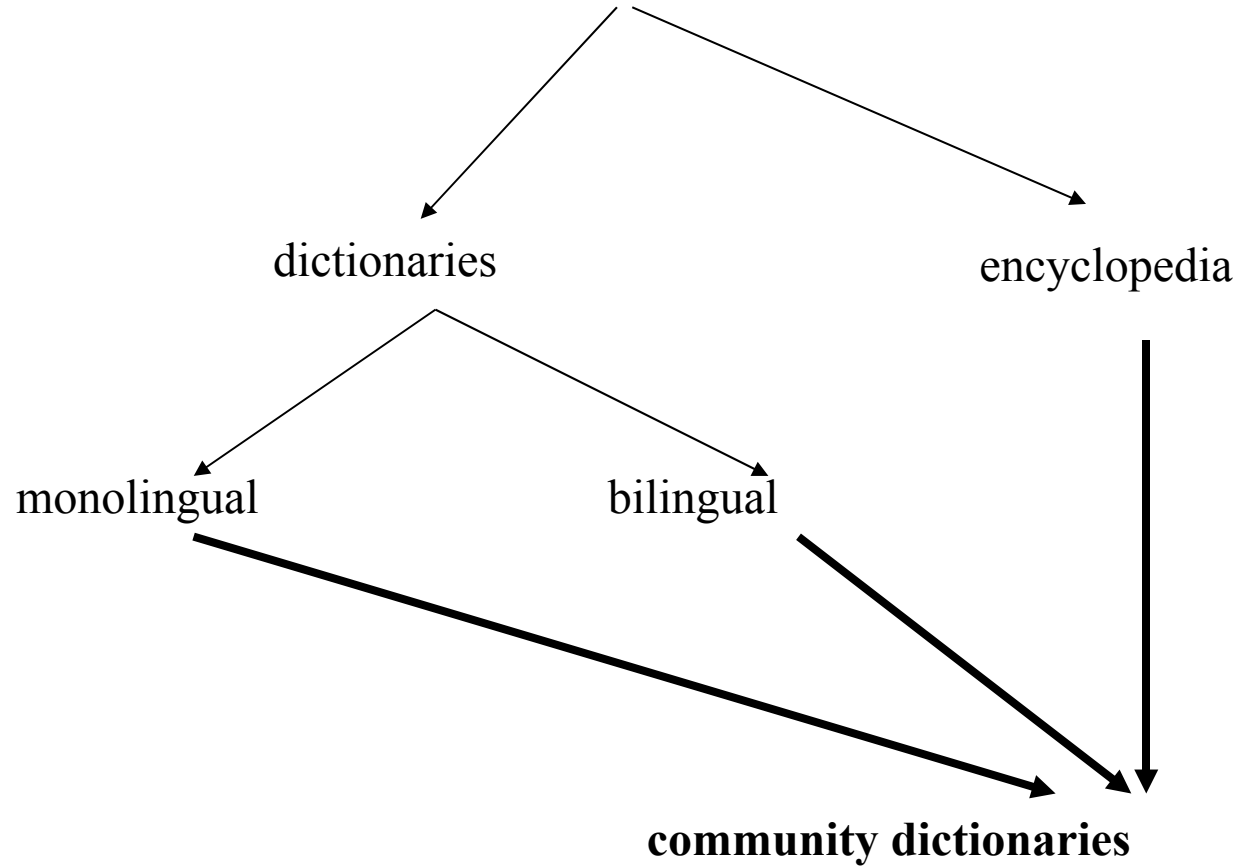
Collection of single sentences
with translation

Photos
drawings
Videos





Typology



Dictionaries for endangered languages are special dictionaries

	ordinary dict.	community dict.
economic basis	commercial	funding agencies
time frame	decades	3-10 years, part time
purpose	translating, L2 learning	language maintenance
users	general public	small community, academics
lexicographers	professionals	linguists, community members
linguistic resources	huge corpora, old dictionaries	language documentation

Dictionaries for endangered speech communities: users & purpose

Linguists

understand and analyse texts

passive

dictionary for translation

linguistic information

Native speakers

Preservation of cultural memory
education

language maintenance,

passive & active

learner dictionary

linguistic &

encyclopedic information

Content and structure of the TLD

Not an end product,
but a **dynamic tool**, containing information on

- the inherent grammatical features of lexical units
(gender, valency)
- grammatical relations between words
(conversion, derivation, composition)

- semantic features
- semantic relations to other lexical units

- translation equivalents
- extralinguistic reality

	Lexical database	Dictionary
Space	unlimited	restricted
Macro- structure	multi-dimensional	linear
Purpose	constantly growing unbiased resource and tool for researchers	user oriented, for the speech community
Content	moderately selective	highly selective
Meanings	text meanings	meaning potentials
examples	citations	illustrative examples

Lexical database

Dictionary

orthography

can easily deal with
variation

standardization preferable

grammar

citations contain
speech errors,
interferences from
dominant languages

standard forms preferable

Problem:
Text meaning vs. meaning potential

Not the meaning of the lexemes captured,
but only the various **senses**
the word has in the particular **contexts** of the corpus.

And what we consider as distinct senses is influenced by the
translation equivalents.


Example: **naovana**

‘and Gaivaa became a **naovana** that we call seagull’

‘then you get two cockatoo feathers, this is a white **naovana**, ...

naovana ‘bird’

but other **naovana**: flying foxes,
insects



“Text meanings arise from combinations,
not from any one word individually.”

Problem:
citations are not suitable
as example sentences.

‘stabbed him ... with her hand’ ?



**‘This old woman,
Sharphand,
stabbed him to death with her hand.’**



the creation of a dictionary is a different job

- revision of sense discrimination
- systematic ordering of senses
- revision of examples

·moderate standardization of orthography
and grammar

The same applies to electronic lexica like LEXUS

You need two versions:

- a) database as a tool for researchers and lexicographers
- b) community version

Problem: time management

The lexical database is too big
to be transformed into a dictionary within a short time..



no comprehensive dictionary
never promise such a dictionary

Necessity is the mother of invention

The first monolingual Samoan dictionary

Ministry of Youth, Culture and Sports, Western Samoa
Australian South Pacific Culture Fund

1 year - 10 000 AUD (6000 €)

compiled 1994
published 1997

How much can be done in one year?

Jakob Grimm, Wilhelm Grimm et al. 1852-1960.
Deutsches Wörterbuch (German Dictionary)
16 volumes

Alphabetical method

A, B, C, D, ...

How do you collect headwords?

start with A

Thematic method

house building,
fishing, ...

- filter the database
- interview experts

Time planning? Setting priorities?

Themes of Teop mini-dictionaries:

house, canoe,
fishes, fishing, the sea, shellfish,
trees, other plants, gardening, cooking,
body, health,
kinship, ceremonies

Further advantages of mini-dictionaries:

specialised vocabulary is less frequently used

- › more endangered
- › less polysemous

**Problem : specialised vocabulary is
under-represented in the corpus**



supplementary recordings

Supplementary recordings in Teop for the House-Dictionary:

How to build a men's house

How to make bamboo walls

How to make the floor from the wood of the *kabuu* palm

How to make the thatch

Further problems of specialised vocabulary

- it is difficult to translate
- linguists have no expertise in ethno-sciences
- indigenous experts lack the necessary proficiency in the target lang.



vernacular encyclopedic descriptions
with translations

Problem: descriptions are not an indigenous genre

- ▶ but show the expressive potential of the language
- ▶ reflect the native speaker's metalinguistic knowledge
- ▶ can show the native speaker's conceptualisation of extra-linguistic phenomena, e.g. taxonomies
- ▶ are useful to understand the meaning of words

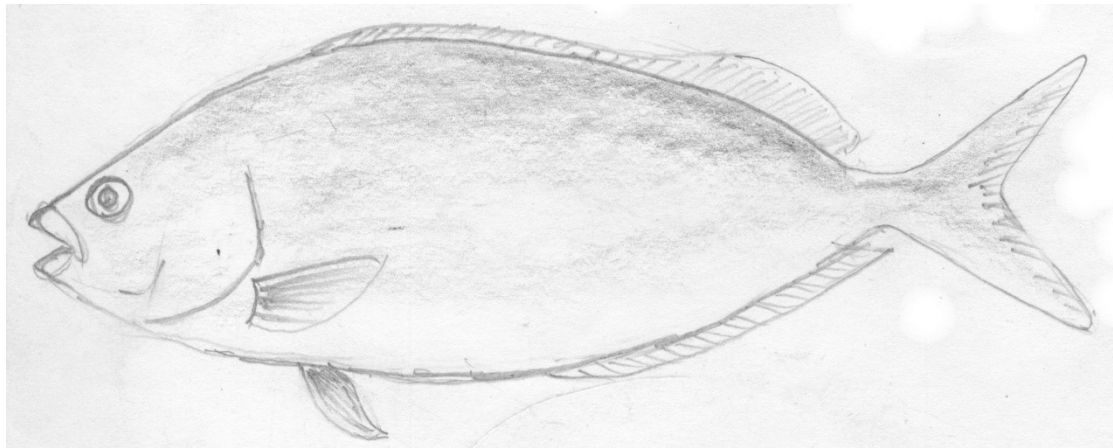


training:

“Explain what X is to a child.”

Sensitive training in writing definitions

marahiri 'The *marahiri* is a fish, that lives in the sea. The *marahiri* has no scales, its body is slippery ' (in Teop language)



What characteristics are essential?

Sensitive training

- avoid patronising
- encourage them to keep and/or develop their own way of explaining words and things
- avoid style guides

naovana ‘bird’

“We eat many birds. Birds are a good food.

Only the birds that have a story are
the ones that we do not eat.

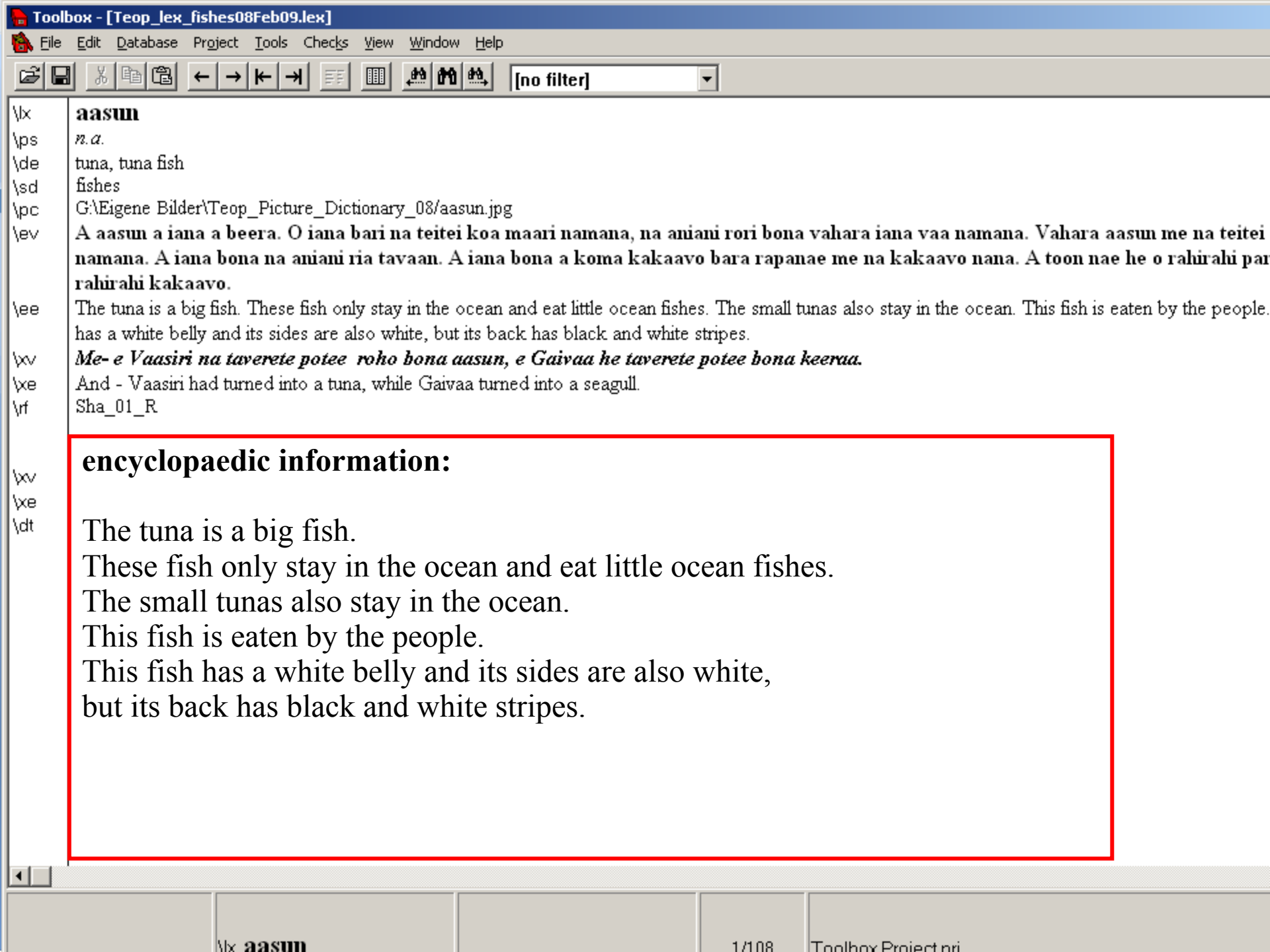
These birds are

Pasukokoreo, Topeipei, Toai und Koo.”

Content of encyclopaedic information:

cf. Coward & Grimes 2000: 138-153)

consult anthropologists, ethnobotanists, etc.



\x **aasun**
\ps *n. a.*
\de tuna, tuna fish
\sd fishes
\pc G:\Eigene Bilder\Teop_Picture_Dictionary_08/aasun.jpg
\ev A aasun a iana a beera. O iana bari na teitei koa maari namana, na aniani rori bona vahara iana vaa namana. Vahara aasun me na teitei namana. A iana bona na aniani ria tavaan. A iana bona a koma kakaavo bara rapanae me na kakaavo nana. A toon nae he o rahirahi par rahirahi kakaavo.
\ee The tuna is a big fish. These fish only stay in the ocean and eat little ocean fishes. The small tunas also stay in the ocean. This fish is eaten by the people. has a white belly and its sides are also white, but its back has black and white stripes.
\vw *Me- e Vaasiri na taverete potee roho bona aasun, e Gaivaa he taverete potee bona keeraa.*
\xe And - Vaasiri had turned into a tuna, while Gaivaa turned into a seagull.
\rf Sha_01_R

\vw
\xe
\dt

encyclopaedic information:

The tuna is a big fish.
These fish only stay in the ocean and eat little ocean fishes.
The small tunas also stay in the ocean.
This fish is eaten by the people.
This fish has a white belly and its sides are also white,
but its back has black and white stripes.

Conventional and idiosyncratic language use in indigenous lexicography

Let different people work on items

that are presumably described in a similar way; e.g.

house & canoe or fishes & birds

similarities of house and canoe descriptions:

topic = thing

predicate = put s.th. somewhere

similarities of fishes and birds descriptions:

topic = animal

predicate = habit

= properties like size, colour, shape

Linguistic observations

conventionalized constructions for descriptions of things, properties and events

systematic patterns of polysemy and word formation

Example for systematic heterosemy/conversion

Noun ‘part of the house’ **Verb** ‘add this part (e.g. wall) onto the house’

‘rafter’

‘put up the rafters’

‘bamboo wall’

‘put up the bamboo wall’

‘fence’

‘put up the fence’

Conclusions

A lexical database and a dictionary are two very different things

a dictionary is not a by-product

a dictionary requires hard work and

active involvement of the speech community



Start with a mini-dictionary.

Use vernacular encyclopedic descriptions.

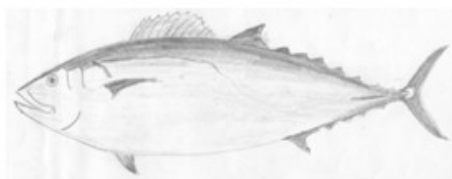
Be rewarded by a culturally and linguistically interesting, and completed little dictionary!

Fish-dictionary

aasun *n.a.* tuna, tuna fish.

A aasun a iana a beera. O iana bari na teitei koa maari namana, na aniani rori bona vahara iana vaa namana. Vahara aasun me na teitei pete rori namana. A iana bona na aniani ria tavaan. A iana bona a koma kakaavo bara rapanae me na kakaavo nana. A toon nae he o rahirahi paru bara rahirahi kakaavo.

The tuna is a big fish. These fish only stay in the ocean and eat little ocean fishes. The small tunas also stay in the Ocean. This fish is eaten by the people. This fish has a white belly and its sides are also white, but its back has black and white stripes. **A bua tom keara paa rahi viravira ni mau bona aasun.** The two brothers caught lots of tuna.



aasun

anahi *n.a.* mullet.

A anahi sa iana beera vira haana,



anahi



o tainana

anahi

araarata *n.a.* coral fish with green and white stripes.

A araarata a iana a monamona. The araarata is a greasy fish. According to Vaabero [Vaa. Sen. "raarata"], araarata and raarata are the same fish.

aranavi *n.a.* red fish found in the ocean and in big rocks in the deep sea.

A aranavi na teitei nana namana bara ruene ei taena ruene ae tea boon. A tabaan teori o vahara kehaa vaa tahii ae tea boon bara o vahara kato me to aniani riori. Aranavi a gogooravi vasihum. A iana bona

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18

References

Atkins, B.T. Sue & Rundell, Michael. 2008.
The Oxford Guide to Practical Lexicography.
OUP.

Coward, David F. & Grimes, Charles E. 2000.
Making Dictionaries,
SIL International Waxhaw. North Carolina.

Hanks, Patrick. 2008. Do word meanings exist? In Thierry Fotenelle.
Practical Lexicography. A Reader. OUP, pp. 125-134

Kilgarriff, Adam. 2008. “I don’t believe in word senses!” In
Thierry Fotenelle.
Practical Lexicography. A Reader. OUP, pp. 135-151